Agric. Sci. Digest., 36 (4) 2016 : 261-266
Print ISSN:0253-150X / Online ISSN:0976-0547

AGRICULTURAL RESEARCH COMMUNICATION CENTRE
www.arccjournals.com

Classification of agricultural productivity index of Cauvery delta zone using artificial neural network

G. Manimannan*[1], C. Arulkumar[2] and R. Lakshmi Priya[3]

Department of Statistics, DRBCCC Hindu College,
Pattabiram, Chennai-60 072, Tamilnadu, India.

Received: 03-03-2016 Accepted: 01-10-2016 DOI:10.18805/asd.v0iOF.4436

ABSTRACT

In this study a novel trial was made to classify the Agriculture Productivity Index (API) for the major crops of Cauvery Delta Zone (CDZ) using neural network and statistical methods. At present the CDZ includes, Tanjavur, Tiruvarur, Nagapattinam, Tiruchirapalli, Pudukottai and Ariyalur districts. The crops grown in the Cauvery delta zone were categorized into four major groups such as, cereals, pulses, oilseeds and cash crops. The data for the period of 2003 to 2012 were collected from the Department of Economics and Statistics, Chennai, Tamilnadu. Enyedi's method was adopted to calculate the API and based on the index the regions were classified by neural network method using Learning Vector Quantization (LVQ). The classification was cross validated statistically, using Multivariate Discriminant Analysis (MDA). The classification results achieved 83% in LVQ and 97% MDA respectively in the entire period of study. The results are obtained as Greater Productivity Regions (GPR), Moderate Productivity Regions (MPR) and Lesser Productivity Regions (LPR) and are plotted in Tamil Nadu spatial map with different colours.

Key words: Agriculture productivity index, Cauvery delta zone, Learning vector quantization, Multivariate discriminant analysis, Novel classification trial, Spatial Pattern.

INTRODUCTION

Agriculture would be an ancient practice of mankind and it would be the first step towards civilization. Agriculture has played a significant role in the development of human society. Indian economy has been considered as an agrarian economy with ¾ of its population in villages having agriculture and allied activities as their primary occupation. In India most of the population gives importance to agriculture because it not only provides food but also the livelihood to more than half of the population. It also supplies lot of goods and raw materials required by the non-agriculture sectors (Sahoo and Sethi, 2012).

Agriculture plays an important role in the economic development process of India and the economic development depends on the net agricultural productivity of an area. The productivity refers to the ratio of index of local agricultural output to the index of total input used in farm production (Sahoo and Sethi, 2012). The Agricultural productivity of an area is indicated by Agricultural productivity index (API). Based on this API the agricultural zones are classified into different categories. This classification of zones by API gains importance for its help in policy making. Among the different methods used for classifying the zone on the basis of API, the Neural Network method is an advanced and recent technique.

The Cauvery Delta Zone (CDZ) is taken for this study. The CDZ lies in the eastern part of Tamil Nadu between 10.00-11.30N latitude and between 78.15 – 79.45 longitudes. It is bordered by the Bay of Bengal on the East and the Palk straight on the South, Trichy district on the west, Perambalur, Ariyalur districts on the North West, Cuddalore district on the North and Puddukkottai district on the South West. At present this zone covers Thanjavur, Nagapattinam, Tiruvarur, Tiruchirapalli, Ariyalur and Pudukkottai districts, which comprises 28 revenue Taluks of the eastern belt of state. All these Taluks are benefited by the river Cauvery. Total area of the zone is 24,943 square kilometers in which 60.2 per cent of the area i.e., 15, 00,680 hectares are under cultivation. Tiruchirappalli, Ariyalur, Pudukkottai, Thanjavur, Thiruvarur and Nagapattinam districts are included under CDZ is known as "Rice Bowl" of Tamilnadu and is depicted in Table 1.

Table 1: Paddy yield ten years period from 2002-2012

Districts	Year	Total Yield of Ten years
Tiruchirappalli	2002-2012	2332208
Ariyalur	2002-2012	275588
Pudukkottai	2002-2012	2169719
Thanjavur	2002-2012	4573543
Thiruvarur	2002-2012	3427869
Nagapattinam	2002-2012	3172246

*Corresponding author's e-mail: manimannang@gmail.com
[1]Department of Statistics, DRBCCC Hindu College, Pattabiram, Chennai-60072, Tamilnadu, India.
[2]Vivekanandha College of Arts and Sciences for Women, Tiruchengode, Namakkal- 637205, Tamilnadu, India.
[3]Department of Statistics, Dr. Ambedkar Government Arts College, Vysarpadi, Chennai 600 039, India.

MATERIALS AND METHODS

Data Collection: The data for the period of 2003 to 2012 were collected from the Department of Economics and Statistics of Tamil Nadu state government, Chennai (Shafi, 1983, Anonymous). The agricultural productivity of 15 major crops growing in this zone was taken for the study. The crops were divided into four categories such as, (i) Cereal crops (which are include paddy, cholam, cumbu, ragi, and maize) (ii) pulses (which are include Bengal gram, red gram, black gram, green gram and horse gram) (c) oil seeds (which are include groundnut, gingerly and coconut) and (d) cash crops (which are include sugarcane and tapioca) and the API was calculated for all these four categories

Calculation of agriculture productivity index: Enyedi's method (Jasbir Singh and Dhillon, 2002) was adopted to calculate the API for CDZ. At present the CDZ includes 6 districts. The agricultural productivity of an area from June to May for the period of 2002-2003 to 2011-2012 was calculated for five districts such as Thanjavur, Nagapattinam, Tiruvarur, Tiruchirapalli and Pudukkottai districts. As the district Ariyalur was originated during the financial year 2009-10, it was excluded from the study, in order to maintain uniformity in comparison. The API is calculated by considering the area of cultivation (Hectare) and production (Tons) of a particular crop in the selected area i.e. district and also in the entire zone i.e CDZ

$$API = \frac{Y}{Y_n} \div \frac{T}{T_n} X100$$

Where, API - Productivity Index

Y – Production of the selected crops in an unit area,

Y_n –Production of the selected crops in the entire zone,

T – Cultivation Area of the selected crops with in the district,

T_n – Cultivation Area of the selected crops with in the entire zone.

Classification of CDZ on API: Based on the API, CDZ was classified by two major techniques like LVQ Neural Network and Multivariate Discriminant Analysis (MDA) by using MATLAB and SPSS software's respectively. The classified zones were labeled as Greater Productivity Regions (GPR), Moderate Productivity Regions (MPR) and Lesser Productivity Regions (LPR) and they were also plotted in Tamil Nadu district map.

Neural network: In machine learning system, there are so many classification methods are available, for example, Support Vector Machine, Quadratic Classification, Expectation and Maximization Algorithm, Random Forest Classification, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), etc., they are categorized by supervised learning and unsupervised learning. In the field neural network, many sophisticated techniques are available for classification, estimation, prediction, etc. these learning methods are closely connected with certain network topology. In figure 1 and 2 some of the both supervised and unsupervised learning methods given in pictorial representation.

Learning vector quantization: In the present study, an attempt is made to classify the total API of various crops in CDZ using LVQ method and also cross validate the success rate of total agriculture productivity index using MDA. It is not only used to solve classification tasks, but also it can be used for simple recognizing. The main purpose of LVQ is to define class regions in the input data space. Learning is based on the definition of quantization areas between neighboring vectors of codebook. This technique is similar to Voronoi sets in vector quantization. Quantization is an approximation of an analog value by one of a finite number of numerical values. It is used to reduce the possible states. The vector quantizer captures a set of vectors into unknown number of final classes, during LVQ training the set is divided into N regions, which are represented by N centroids. A set of code vectors (centroids) structure the codebook Kohonen T [4]. Vector quantization is a standard statistical clustering

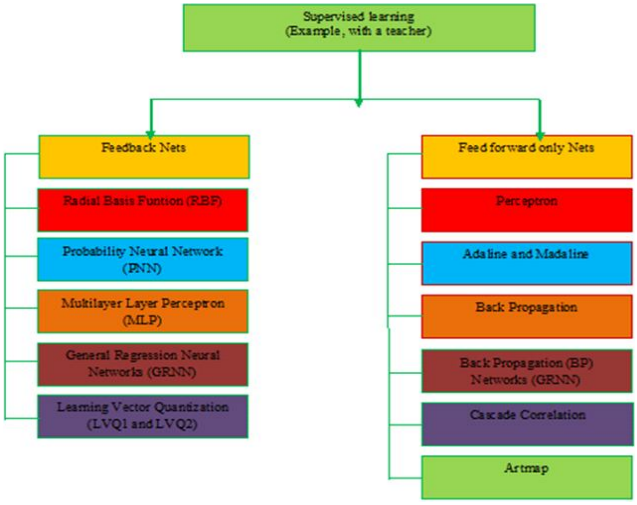

Fig 1: Methods of unsupervised learning neural network

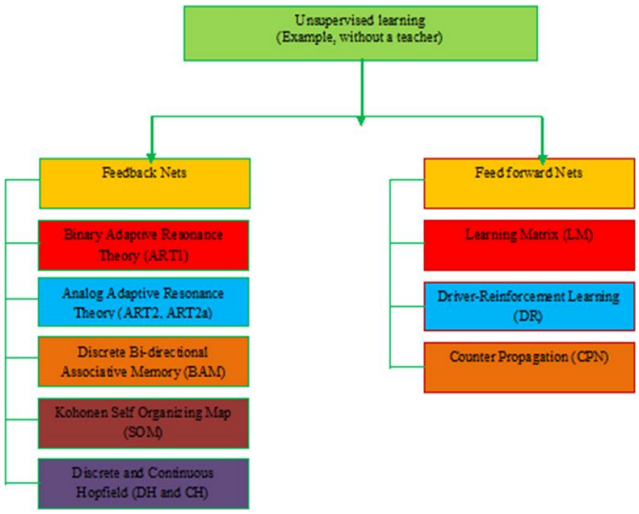

Fig 2: Methods of Supervised learning neural network

technique, which seeks to divide the input space into areas that are assigned as *code book* vectors.

Standard algorithm for learning vector quantization: The typical Kohenen network used for training pattern and classification. This LVQ network holds information about association in the classes. This attains coarse adjustment of vectors to probable future clusters. Then to all vectors is allocates symbol of the class which represents class according to the nearest centroid.

In addition, the location of classes for the classification of a new pattern is adjusted and if the pattern cannot be assigned to an existing class, a new class is formed.

Learning juncture is slightly more complicated and alternates of this learning are LVQ1, LVQ2 and LVQ3, which differ in their search for the optimal boundaries between classes.

LVQ clustering algorithm: The LVQ1 training process proceeds with an input vector randomly selected from the labeled training set. Given an input vector u_i to the network, the output neurons in LVQ1 are deemed to be a *winner* according to the following.

$$\min d(u_i, w_j) = \min \left\| u_i - w_j \right\|^2$$

Let $\{u_i\}$, for $i = 1,2,..., N$ be the set of input vectors, and the network synaptic weight vectors are denoted by $\{w_j\}$for j= 1,2,.....m. The weight vector w_j is adjusted in the following manner:

1. If the class associated with the weight vector and the class label of the input are the same, that is $Cw_j = Cu_i$, then $w_j(k+1) = w_j(k) + \mu(k)[\mu_i - w_j(k)]$, where $0 < \mu(k) < 1$ (the learning rate parameter)
2. If $Cw_j = Cu_i$ then $w_j(k+1) = w_j(k) + \mu(k)[\mu_i - w_j(k)]$ and other weight vector not adopted.

Therefore, the update rule for modifying a weight vector in (Anonymous, 1984) is standard one if the class is correct. In other words, according to the learning rule (1), the weight vector w_j, is moved in the direction of the input μ_i, if the class labels of the input vector and the weight vector agree. In case, the class is not correct, the weight vector is moved in the opposite direction away from the input vector according to the rule (Jasbir Singh and Dhillon, 2002). The learning rate parameter $\mu(k)$ is monotonically decreased in accordance with the discrete index *k*. Commonly used learning rate initial value 0.1 and it is linearly decreased. The stopping condition based on the total number of desired training epochs or monitoring the convergence of the weight vectors. Another stopping condition based on monitoring the learning rate parameter directly, when it is sufficiently small training can be terminated. In this research paper, we established the stopping condition to be the total number (predefined) of iterations. The basic algorithm can be summarized as follows:

Step 1: Initialize all weight vectors $w_j(0)$, initialize the learning rate parameter $\mu(0)$, and set $k = (0)$.
Step 2: Check the stopping condition. If the condition is false, continue; if the condition is true then quit.
Step 3: For each training vectors μ_i, perform step 4 and 5.
Step 4: Determine the weight vector index ($j = q$) such that min $\| \mu_i - w_j \|^2$ and the weight vector $w_q(k)$ that minimizes the square of the norm.
Step 5: Update the appropriate weight vector as follows: If cc,

$$Cw_q \neq Cu, then\ w_q(k+1) = w_q(k) - \mu(k)[u_i - w_q(k)]$$

Step 6: Set, reduce the learning rate parameter, then go to step 2. The learning rate parameter can be reduced in accordance with (discrete-time index) using

$$\mu(k+1)/(k+1) \quad for\ k > 0$$

The neural network architecture for LVQ1 is shown in figure (3):

Multivariate discriminant analysis: Multivariate discriminant analysis is statistical analysis which is used to derive the linear combination of two or more independent variables that will discriminate best between a priori defined groups, which is generally, male or female, failure or non-failure etc. This is achieved by the statistical decision rule which maximizes the between group variance relative to the within group variances. The multiple discriminant analysis derives the linear combination from an equation that takes the following form

$$Z = W_1 X_1 + W_2 X_2 + ... + W_n X_n$$

When Z discriminant score, W_i – Discriminant weight and X_i – Independent variables.

In the present study, MDA is applied to the agriculture productivity indices to cross validate the classification obtained in [1].

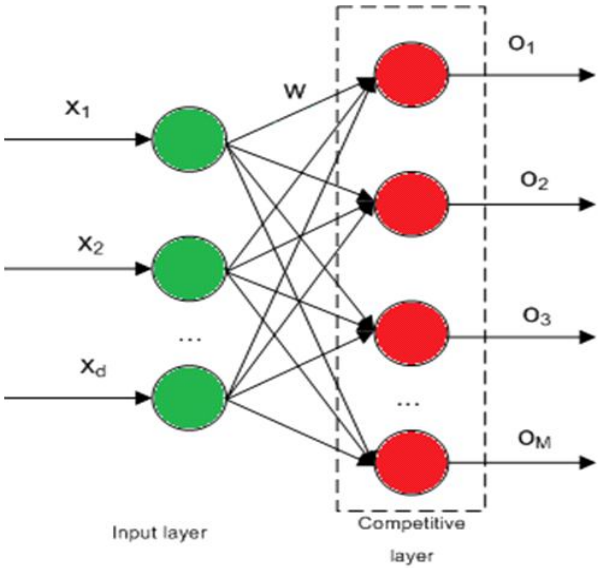

Fig 3: Neural network architecture for LVQ1

RESULTS AND DISCUSSION

In the present study an attempt is made to find the agriculture productivity indices and classification of agricultural productivity indices using LVQ neural network and also cross validate with MDA during the period 2003 to 2012. Initially Enyedi's method was adopted to compute Agriculture productivity index. The researcher have to associated fifteen samples of agriculture productivity indices (API) (Pulses, cereals, oil seeds and cash crops parameters) to class 1, twenty one samples of API to class 2 and eighteen samples of API to class 3 and thereby we have totally 53 calculated API vectors associated with three class.

Method I: To initialize the codebook vector w_1, w_2 and w_3 is to use the first three samples, one from each group of samples of each class and labeled as 1, 2 and 3 respectively. The remaining 51 samples of the three classes can be used for training. We initialize the learning parameter to and decrease it by every training epoch. According to the LVQ clustering algorithm stated above, we set k=2, μ (2=μ (1)/ 2=0.5, then go to step 2 and check the stopping condition. The next training epoch we start with the first training sample among the remaining 20000 API of three classes. The MATLAB coding is used the number of training epochs set to 20000 to find *codebook* vectors. After training 20000 training epochs, the final weights were given in Table 1 and also found actual classification as well estimated classification for API (Table 2)

Method II: For cross validate the results the same API are used in MDA. In all the districts of CDZ are classified during the training process just 70 epochs were performed using API. There has been 83.0% success rate attained. In the both neural network training epochs and MDA (20000 and 70) the LVQ achieved same classification rates. API of 38 samples were correctly classified from the period of 2003 to 2012 and 15 cases were misclassified during the year 2004, 2005 and 2006 in the case of MDA. The codebook vectors and summary statistics are given in Table 2 and 3. Finally, the total classified results of LVQ and MDA using the index and labeled as LPR, MPR and HPR in Table 3.

In the Table 2, results of LVQ algorithm calculated in MATLAB and results of MDA method calculated in SPSS are compared. It can be seen that healthier results were obtained with LVQ and MDA algorithm. Only in three years (2004, 2005 and 2006) in the LVQ trained algorithm was misclassified than MDA method with 0.3 % variation. The remaining case in LVQ shows better results in CDZ. The overall classification achieved by MDA (97 %) and LVQ (83.0%) for total API. For the training of this attribute 20000 and 70 epochs has been used. The classification success rate reached 83 percentages. The values were classified into three classes which are shown in figure 4. In the figure 4 on the left side it can be seen that in the beginning the error decreased then it started to swing and eventually stabilize during training for 70 epochs.

Figure 4, shows that LVQ best training map of the overall data matrix and in figure 4 result of LVQ Receiver Operating Characteristics (ROC) curve for the training time of the entire API from the period of 2003-2012. Figure 4, shows that the results of overall performance of API confusion matrix of API index data. In figure 7 shows that the results of scatter diagram of overall neural network classification. Finally, the researcher highlighted the indices in the spatial map of CDZ (Figure 5) based on simulation of LVQ classification. It is observed that from Fig. 5 the total API fluctuated in every year in the district of CDZ using LVQ and MDA. In the year 2003 all districts belongs to MPR category except the district Nagapattinam fit in GPR. In 2004, Trichy has GPR, Tiruvarur in LPR category and remaining districts in MPR. Two districts (Trichy and Nagapattinam) are belonging to GPR and rest of the district in MPR for the year 2005. In the year of 2006 and 2008, Nagapattinam in GPR, Thanjavur is in MPR and remaining districts in LPR. Nagapattinam, Trichy are observed in HPR category and the remaining districts falls under MPR in the year 2007.

Very poor performance in the API has been observed in the years 2008, 2009, 2010, 2011, 2012, because all the districts belonging to MPR and LPR categories except in year 2009. Interestingly, the spatial mapping results showed that the rice bowl of Tamilnadu, Thanjavur district observed in the category of MPR except 2009 (GPR) and 2010(LPR). Various reasons to quote for fluctuations in the API in CDZ, and they are Cauvery water dispute, insufficient rainfall for better irrigation, urbanization, booming corporate

Table 2: Codebook of four major crops (Training of 20000 epochs)

Crops Categories	Class 1	Class2	Class 3
Cereals	92.4073	65.4424	127.8306
Pulses	33.3061	138.3391	216.4931
Oil Seeds	82.2460	90.2552	115.9631
Cash Crops	10.1846	180.8641	67.1039

Table 3: Summary Statistics of classification of LVQ and cross Validation of MDA for API of cauvery delta zone (2003-2012)

Year	LVQ Classification			Total	MDA Classification			Total
	LPR	MPR	HPR		LPR	MPR	HPR	
2003	0	4	1	5	0	4	1	5
2004	1	3	1	5	0	4	1	5
2005	0	3	2	5	1	2	2	5
2006	3	1	1	5	2	2	1	5
2007	1	3	1	5	1	3	1	5
2008	3	1	1	5	3	1	1	5
2009	3	2	0	5	3	2	0	5
2010	2	4	0	6	2	4	0	6
2011	3	2	1	6	3	2	1	6
2012	3	3	0	6	3	3	0	6
Total	19	26	08	53	18	27	08	53

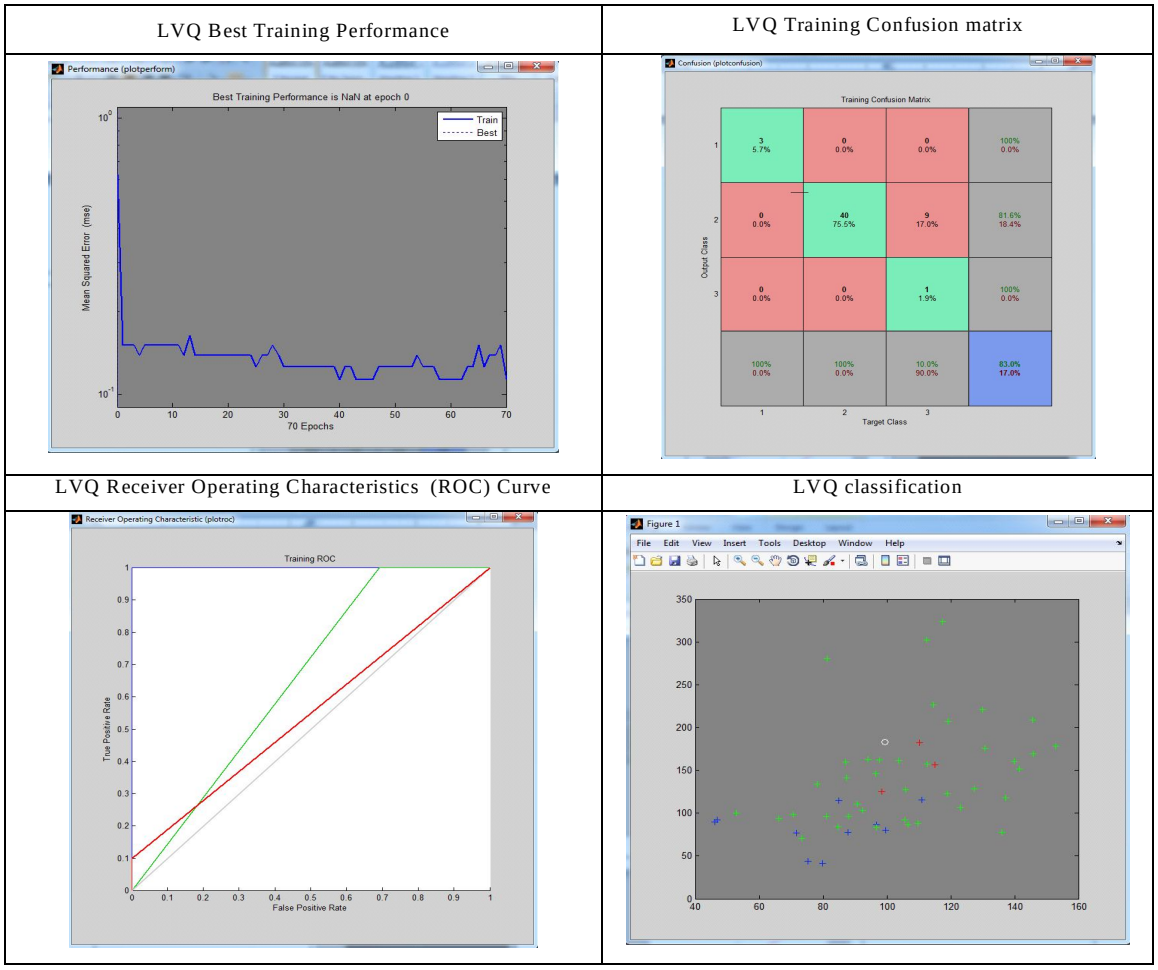

Fig 4: Agriculture Productivity Index for LVQ Neural Network classification (2003-2012)

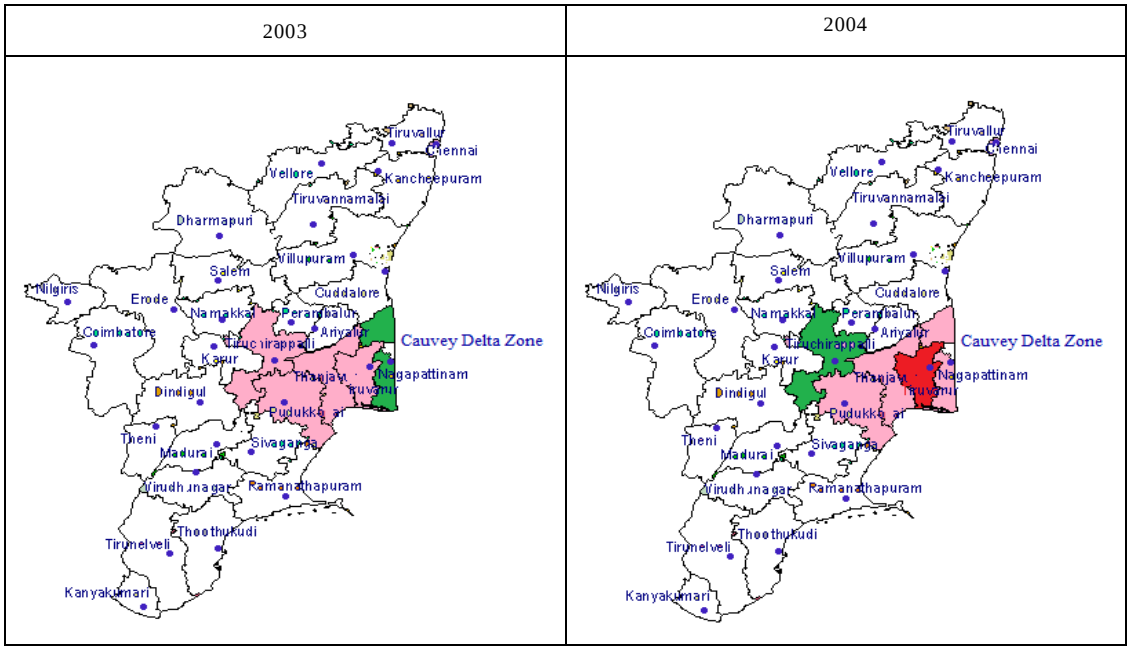

Fig 5: Sample Spatial Pattern of Agriculture Productivity Index Based on LVQ Classification (2003-2004)

companies particularly in delta districts. In 2008, Cyclone Nisha crushed the districts of Cauvery delta zone. The development discuss in the Cauvery delta districts faces another challenge in a project proposed to extract methane from lignite seams over 667 sq km, in the UPA government at the Centre gave the approval for a 'Coal Bed Methane (CBM)' pilot project to be executed in the 'Mannargudi block' by a Gurgaon company (Source: The Hindu June16, 2014).

CONCLUSIONS

This study focused on estimation of API and classification based on the API by using two methods of LVQ for classification and MDA for cross validation. From the performed simulation and statistical experiments over the given data, it is clear that for a given application is crucial to use a suitable type of artificial neural networks and statistical methods. Classification results obtained by means of Fisher linear discriminant function and use of LVQ neural network in the Cauvery Delta Zone of Tamilnadu. Currently, our research focuses on testing different topologies of neural networks LVQ and also other types of neural networks in the Agriculture data set. In both methods, the classification achieved 83 % in LVQ and 97 % in MDA respectively in the ten years period of study. In the ten year periods agriculture productivity index data fluctuated in all the years in all the districts of CDZ, because the main reasons are rainfall fluctuation, uncertainity dates of opening and closing of Cauvery water, undelayed monsoons, real estate business *etc*., Another reason of districts classification depends on various types of soils. Finally, our research concludes three types of classification achieved using MDA and LVQ neural networks during the period of 2003-2012 in the area of CDZ. Agriculture Productivity indices labeled as LPR, MPR, GPR based on over all means and are plotted in the form of spatial map in each year of the study period. A generalization of the results is under investigation to obtain a set of 3 groups of API for any given year.

REFERENCES

Anderson TW (1984) An Introduction to Multivariate Statistical Analysis, 2/e, John Wiley and Sons, Inc., New York.

Jasbir Singh, Dhillon S S (2002) Agricultural Geography, Tata McGraw Hill Publishing Company Limited, New Delhi.

Kalpana Sahoo, Narayan Sethi (2012) Investigating the Impact of Agriculture and Industrial Sector on Economic Growth of India

Industrial Sector on Economic Growth of India. OIDA International Journal of Sustainable Development 05:05, pp. 11-22.

Kohonen T (2001) Learning vector quantization. Springer Berlin Heidelberg.

Shafi M (1983) Agricultural Productivity and Regional Imbalances – A study of Uttar Pradesh. Concept Publishing Company, New Delhi.

Anonymous, (2016). Season and Crop Report, Tamilnadu (Various Issues), published by Department of Economics and Statistics, Chennai.

Anonumous (2012-13). Statistical Hand Book of Tamilnadu (2012-2013), Department of Economics and Statistics, Chennai. (http://agritech.tnau.ac.in/pdf/stst_handbook_tn_2013.pdf)

http://www.tnau.ac.in/dr/zonepdf/CauveryDeltaZone.pdf